2/9/1
DIALOG(R)File 351:Derwent WPI

011100617 **Image available**
WPI Acc No: 1997-078542/*199708*
XRPX Acc No: N97-065161

Input parameter preparation for neural network - determining missing measurement value from values adjacent in time series and from determined statistical noise distribution, and substituting value with at least two Monte-Carlo draws, based on fault value noise distribution

Patent Assignee: SIEMENS AG (SIEI)
Inventor: TRESP V
Number of Countries: 003 Number of Patents: 003
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
DE 19530647	C1	19970123	DE 1030647	A	19950821	199708	B
JP 9128361	A	19970516	JP 96217589	A	19960819	199730	
US 5706401	A	19980106	US 96699558	A	19960819	199808	

Priority Applications (No Type Date): DE 1030647 A 19950821
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
DE 19530647	C1		11	G06F-017/18	
JP 9128361	A		9	G06F-015/18	
US 5706401	A		8	G06E-001/00	

Abstract (Basic): DE 19530647 C

The method comprises the steps of forming a time series from a number of measurement values of a variable input parameter, by imputing values at discrete time intervals, and supplying and/or calculating a statistical noise distribution from the unrelated noise of limited variation, that overlaps the input signal, and that has a mean value of zero.

A missing measurement is entered into the time series as a fault value, by calculating a statistical fault value from at least one value, adjacent in the time series, and from the determined statistical noise distribution. At least two Monte-Carlo draws are determined from the fault value noise distribution, and the resulting value is entered in place of the fault value.

USE/ADVANTAGE - For training of neural network. Improves quality of sequential input values of network. Substitutes missing parameters, or parameters lost in noise.

Dwg.2/3

Abstract (Equivalent): US 5706401 A
The method comprises the steps of forming a time series from a number of measurement values of a variable input parameter, by imputing values at discrete time intervals, and supplying and/or calculating a statistical noise distribution from the unrelated noise of limited variation, that overlaps the input signal, and that has a mean value of zero.

A missing measurement is entered into the time series as a fault value, by calculating a statistical fault value from at least one value, adjacent in the time series, and from the determined statistical noise distribution. At least two Monte-Carlo draws are determined from the fault value noise distribution, and the resulting value is entered in place of the fault value.

USE/ADVANTAGE - For training of neural network. Improves quality of sequential input values of network. Substitutes missing parameters, or parameters lost in noise.

Dwg.2/3

Title Terms: INPUT; PARAMETER; PREPARATION; NEURAL; NETWORK; DETERMINE; MISS; MEASURE; VALUE; VALUE; ADJACENT; TIME; SERIES; DETERMINE; STATISTICAL; NOISE; DISTRIBUTE; SUBSTITUTE; VALUE; TWO; DRAW; BASED; FAULT; VALUE; NOISE; DISTRIBUTE
Derwent Class: T01; T06
International Patent Class (Main): G06E-001/00; G06F-015/18; G06F-017/18
International Patent Class (Additional): G05B-013/02; G06E-003/00; G06F-017/17
File Segment: EPI
Manual Codes (EPI/S-X): T01-J16C1; T01-J16C2; T06-A05

98 P 5929

(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) **Pat ntschrift**

(10) **DE 195 30 647 C 1** B 7

(51) Int. Cl.⁶:
G 06 F 17/18
G 05 B 13/02
G 06 F 17/17

(21) Aktenzeichen: 195 30 647.3-53
(22) Anmeldetag: 21. 8. 95
(43) Offenlegungstag: —
(45) Veröffentlichungstag der Patenterteilung: 23. 1. 97

DE 195 30 647 C 1

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

(73) Patentinhaber:
Siemens AG, 80333 München, DE

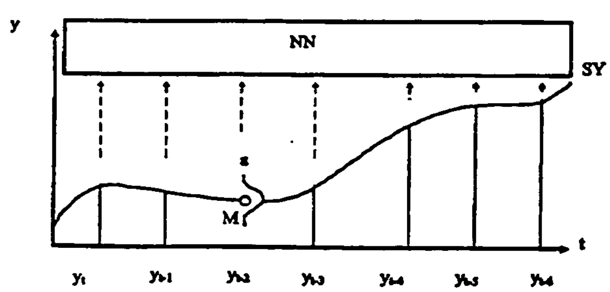
(72) Erfinder:
Tresp, Volker Dr., 80539 München, DE

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:

DE	42 41 812 A1
DD	2 66 661 A1
EP	05 04 457 A1

(54) Verfahren zur Aufbereitung einer Eingangsgröße für ein neuronales Netz

(57) Durch das erfindungsgemäße Verfahren werden Wege aufgezeigt, fehlende Daten in Zeitreihen von neuronalen Netzen zu ergänzen, oder verrauschte Daten zu verbessern. Hierzu werden anhand der bekannten Meßwerte aus der Zeitreihe und deren bekannter oder vorgegebener Fehlerverteilungsdichte für die fehlenden Werte Fehlerverteilungsdichten berechnet und gemäß der Monte Carlo Methode aus dieser Fehlerverteilungsdichte Proben gezogen, die zu je einem prognostizierten Wert führen, deren Mittelwert für den zu prognostizierenden Wert eingesetzt wird. Das Verfahren kann dabei sowohl für den Betrieb, als auch für das Training der Netze verwendet werden. Einsatzgebiete der Erfindung liegen auf allen bekannten Einsatzgebieten neuronaler Netze.



DE 195 30 647 C 1

Beschreibung

Die Erfindung bezieht sich auf ein Verfahren zur Aufbereitung einer Eingangsgröße eines neuronalen Netzes, d. h. zur neuronalen Modellierung von dynamischen Prozessen, von denen insbesondere unv llständige oder schlechte Meßreihen vorliegen.

Neuronale Netze finden in die vielfältigsten technischen Gebiete Eingang. Überall dort, w es gilt, aus komplexen technischen Zusammenhängen und aus unzureichenden Informationen Entscheidungen abzuleiten, erw isen sich neuronale Netze als besonders geeignet. Zur Bildung iner oder mehrerer Ausgangsgrößen werden dem neuronalen Netz beispielsweise eine oder mehrere Eingangsgrößen zugeführt. Hierzu wird ein solches Netz zunächst für den speziellen Einsatzfall trainiert. Neuronale Netze erweisen sich für viele Einsatzfälle als besonders geeignet, da sie universelle Approximatoren sind.

Ein häufig auftretendes Problem im Zusammenhang mit dem Einsatz von neuronalen Netzen besteht allerdings darin, daß häufig die Eingangsdaten zum Training, oder beim Betrieb des Netzes nicht vollständig sind. Dieser Sachverhalt und auch die Tatsache, daß die Meßwerte für den Aufbau einer Zeitreihe, welche dem neuronalen Netz zugeführt wird häufig ungenau oder verrauscht sind, bewirken, daß teilweise schlecht Lernergebnisse der Netze erzielt werden.

Aus dem Stand der Technik sind verschiedene Aufbereitungsmöglichkeiten für Eingangsgrößen bekannt. In der Patentschrift DD 2 66 661 A1 ist eine Schaltungsanordnung zur Approximation von Meßwertfolgen beschrieben, bei der das Problem der Datenverarbeitung von verrauschten Meßwerten in Echtzeit dadurch gelöst wird, daß eine Approximation über die Schätzung der Zustandsgrößen erfolgt. Unbekannte Größen werden dabei nach dem Prinzip der maximum-likelihood ermittelt. Aus der europäischen Patentanmeldung EP 05 04 457 A1 ist ein Verfahren zur Ermittlung der Wahrscheinlichkeiten fehlender Betriebszustände bei technischen und biologischen Systemen unter Verwendung von Einzelereigniswahrscheinlichkeiten bekannt. Dort werden die Wahrscheinlichkeiten fehlerhafter Betriebszustände eines technischen oder biologischen Systems dadurch ermittelt, daß ein allgemeiner Interpolationssatz für die Konstitution einer Vielzahl von Interpolationsformeln für Wahrscheinlichkeiten einer dort angegebenen bestimmten Form angegeben wird. Aus der deutschen Offenlegungsschrift DE 42 41 812 A1 ist ein Verfahren zur adaptiven Quantisierung eines Eingangswertebereiches bekannt. Zur adaptiven Quantisierung eines Eingangswertebereiches von Eingangswerten eines Rechners wird dort die Anwendungen eines Histogrammdichteschätzverfahrens vorgeschlagen. Dabei werden die Eingangswerte eingelesen und einer Quantisierung des Eingangswertebereiches zugeordnet, wonach aus der Verteilung der Eingangswerte eine Dichteschätzwertfunktion errechnet wird, welche in eine monoton steigende Abbildungsfunktion transformiert wird.

Andere Möglichkeiten der Aufbereitung von Eingangsgrößen sind derzeit aus dem Stand der Technik nicht bekannt.

Die der Erfindung zugrundeliegende Aufgabe besteht darin, ein Verfahren anzugeben, mit dem die Qualität von zeitlich nacheinander anfallenden Eingangsgrößen eines neuronalen Netzes verbessert werden kann. Insbesondere sollen durch das erfindungsgemäße Verfahren fehlende und verrauschte Daten ergänzt werden.

Diese Aufgabe wird gemäß den Merkmalen des Patentanspruchs 1 gelöst.

Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

Ein besonderer Vorteil des erfindungsgemäßen Verfahrens besteht darin, daß ausgenutzt wird, daß die fehlenden Werte oder die verrauschten Werte, welche den neuronalen Netz zugeführt werden sollen, Bestandteil einer Abfolge von Werten in der Zeitreihe sind. Vorteilhaft kann die bekannte Fehlerverteilungswahrscheinlichkeit der restlichen Werte dazu benutzt werden, um für den fehlenden Wert nach dem erfindungsgemäßen Verfahren eine erwartete Fehlerverteilung und damit den erwarteten Wert berechnen zu können.

Günstigerweise können nach dem erfindungsgemäßen Verfahren auch fehlende Werte, die in der Zeitreihe benachbart sind bestimmt werden. Dafür ist ein iterativer Vorgang vorgesehen, der einmal den einen Wert berechnet und danach den anderen mit den aus dem einen Wert gewonnenen Daten ermittelt. Günstigerweise kann dieser Iterationsvorgang auch mehrfach durchgeführt werden, damit eine hinreichende Genauigkeit der zu bestimmenden Werte gewährleistet ist.

Besonders kann nach dem erfindungsgemäßen Verfahren auch ein neuronales Netz trainiert werden, welches die Zeitreihe nachbilden soll, da dabei vorteilhaft die Lernschrittweite auf die Zahl der gezogenen Monte Carl Proben bezogen wird.

Günstigerweise wird für das Fehlerverhalten der vorhergesagten Werte der Zeitreihe eine Gaußverteilung angenommen, oder festgelegt da dies eine Verteilung ist, welche praxisnahen Werten weitestgehend entspricht.

Vorteilhaft werden einfache numerische Verfahren beim erfindungsgemäßen Verfahren eingesetzt, um fehlende Werte bestimmen zu können bzw. um einen künftigen Wert in der Zeitreihe mit fehlenden Werten vorhersagen zu können.

Vorteilhaft werden nach dem erfindungsgemäßen Verfahren einfache mathematische Methoden angegeben, um Meßwerte entsprechend aufbereiten zu können und damit das neuronale Netz trainieren zu können.

Besonders vorteilhaft werden Methoden angegeben, um verrauschte Meßwerte aufzubereiten, bzw. um Meßwerte aufzubereiten, welche eine bekannte und eine unbekannte Rauschkomponente enthalten, um damit das neuronale Netz auf einfache Weise und effizient trainieren zu können.

Im folgenden wird die Erfindung anhand v n Figuren weiter erläutert.

Fig. 1 zeigt ein Z itreihe,

Fig. 2 zeigt eine Zeitreihe und ein Systemverhalten,

Fig. 3 zeigt in neuronales Netz welches traini rt wird.

Fig. 1 zeigt eine Zeitreihe von Meßwerten, welche beispielsweise einem neuronalen Netz zugeführt werden können. Gemäß ihrer zeitlichen Abfolge werden diese Meßwerte beispielsweise von einem technischen System

erfaßt und gemäß ihrer zeitlichen Abfolge mit y_t bis y_{t-6} bezeichnet. Die dargestellten Pfeile zwischen den einzelnen Kästchen symbolisieren die Abhängigkeiten der verschiedenen Werte untereinander. Beispielsweise wird in Fig. 1 dav n ausgegangen, daß der Wert y_{t-2} fehlt. Die im Markov blanket relevanten Werte, als benachbarte Werte dieses fehlenden Meßwertes, sind y_{t-4}, y_{t-3}, y_{t-1} und y_t. Ein solch fehlender Meßwert in einer Zeitreihe kann beispielsweise dadurch entstehen, daß zum fraglichen Zeitpunkt das Meßgerät zur Werteaufnahme nicht funktionierte, oder daß es zwischen einzelnen gemessenen Werten günstig erscheint, um das neuronale Netz besser zu trainieren, diesem einen weiteren Wert zuzuführen, der folglich noch zu bestimmen ist, also nach dem erfindungsgemäßen Verfahren erzeugt werden soll.

Fig. 2 zeigt Zeitreihe aus Fig. 1 in Verbindung mit einem neuronalen Netz NN. Es ist zu erkennen, daß y eine zeitabhängige Variable darstellt, welche das Systemverhalten SY eines technischen Systems repräsentiert. Wie erkannt werden kann, entsprechen die Werte y_t bis y_{t-6} Meßwerten, welche dem Systemverlauf SY entnommen werden. Durch die gestrichelten Pfeile zu den jeweiligen Zeitpunkten ist symbolisiert, daß diese Meßwerte dem neuronalen Netz NN beim Betrieb oder beim Training zugeführt werden sollen.

Wie auch in Fig. 1 ist der fragliche Meßwert M für den Zeitpunkt y_{t-2} nicht vorhanden. Für diesen Meßwert M ist seine Wahrscheinlichkeitsdichte ε angegeben. Diese Wahrscheinlichkeitsdichte ε kann beispielsweise nach dem erfindungsgemäßen Verfahren aus einer vorgegebenen bekannten Fehlerverteilungsdichte der übrigen Meßwerte rückgerechnet werden. Insbesondere wird dabei ausgenutzt, daß sich der fehlende Meßwert zwischen zwei bekannten Meßwerten befinden muß und damit auch dessen Fehler durch die Fehler der benachbarten und der restlichen Meßwerte der Zeitreihe begrenzt wird. Die zugrundeliegende Zeitreihe läßt sich wie folgt beschreiben:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-N}) + \varepsilon_t \qquad (1)$$

Dabei ist f entweder bekannt oder wird hinreichend durch ein neuronales Netz modelliert. ε_t bedeutet dabei einen additiven unkorrelierten Fehler mit zeitlichem Mittelwert 0. Dieser Fehler weist dabei und das ist für das erfindungsgemäße Verfahren essentiell eine bekannte oder vorgegebene Wahrscheinlichkeitsdichte $P_\varepsilon(\varepsilon)$ auf und versinnbildlicht typischerweise die unmodellierte Dynamik der Zeitreihe. Beispielsweise soll für eine solche Zeitreihe, die nach dem erfindungsgemäßen Verfahren komplettiert werden soll, ein zukünftiger Wert vorhergesagt werden. Dabei ist zu beachten, daß zukünftige Werte relativ zu der momentanen gewählten Zeitposition zu verstehen sind. Das heißt für einen Zeitpunkt y_{t-5} ist der Zeitpunkt y_{t-4} ein zukünftiger Wert. Unter diesen Voraussetzungen läßt sich die bedingte Wahrscheinlichkeitsdichte für einen vorherzusagenden Wert der Zeitreihe wie folgt beschreiben.

$$P(y_t | y_{t-1}, y_{t-2}, \ldots, y_{t-N}) = P_\varepsilon(y - (y_{t-1}, y_{t-2}, \ldots, y_{t-N})) \qquad (2)$$

Wie bereits erwähnt muß die Fehlerverteilungsdichte bekannt sein. Diese Verteilungsdichte kann entweder anhand des Systemverhaltens und bekannter anderer äußerer Größen ermittelt oder vorgegeben werden. Eine typische Fehlerverteilung, die in der Praxis auftritt ist die Gaußverteilung. Mit einer solchen angenommenen Gauß'schen Fehlerverteilung läßt sich die bedingte Wahrscheinlichkeitsdichte wie folgt beschreiben:

$$P(y_t | y_{t-1}, y_{t-2}, \ldots, y_{t-N}) = G(y_t; f(y_{t-1}, \ldots, y_{t-N}), \sigma^2) \qquad (3)$$

Darin bedeutet $G(x; c, \sigma^2)$ die Notation für eine normale Dichte, die bei x bestimmt wird mit einem Zentrum C und einer Varianz σ^2. Geht man davon aus, daß das zu beschreibende System in Form einer Folge von Werten auf einer Zeitachse dargestellt wird, so kann man die einzelnen Werte von y_t auch als Zufallsvariable in einem probabilistischen Netzwerk auffassen. Dabei liegt der Erfindung das Problem zugrunde, einen Wert der Zeitreihe vorherzusagen, indem die vorhandene Information aus den restlichen Werten möglichst vollständig verwendet wird. Unter Voraussetzung der Annahmen, die zuvor gemacht wurden, läßt sich die gesamte Wahrscheinlichkeitsdichte der Zeitreihe wie folgt beschreiben:

$$P(y_1, y_2, \ldots, y_t) = P(y_1, \ldots, y_N) \prod_{l=N+1}^{t} P(y_l | y_{l-1}, \ldots, y_{l-N}) \qquad (4)$$

Dabei wird davon ausgegangen, daß y_{t-k} mit $k \leq N$ der fehlende Wert ist. Vorausgesetzt, daß $y^u = \{y_{t-k}\}$ und $y^m = \{y_{t-1}, \ldots, y_{t-k-N}\} \backslash \{y_{t-k}\}$ gilt, kann der erwartete Wert der in der Zeitreihe vorherzusagen ist wie folgt beschrieben werden:

$$E(y_t | M_{t-1}) = \int f(y_{t-1}, \ldots, y_{t-k}, \ldots, y_{t-N}) P(y^u | y^m) dy^u \qquad (5)$$

Dabei gelten folgende Voraussetzungen:
M_{t-1} steht für alle Messungen bis zum Zeitpunkt $t-1$. Die voranstehende Gleichung ist die grundlegende Gleichung für die Vorhersage mit fehlenden Daten. Dabei ist besonders zu beachten, daß die Unbekannte y_{t-k} nicht nur von den Werten der Zeitreihe vor dem Zeitpunkt $t-k$ abhängt, sondern auch von den Messungen nach t-k. Der Grund besteht darin, daß di Variablen in $y^m \cup y_t$ ein minimales Markov blanket von y_{t-k} formen. Dieses minimale Markov blanket besteht aus den direkt n Vorfahren und den direkten Nachfahr n einer Variable und allen direkten Vorfahren von Variablen des direkt n Nachfolgers. Im betrachteten Beispiel in

Fig. 2 sind die direkten Nachfahren $y_t \ldots y_{t-k+1}$. Die direkten V rfahren sind:

$y_{t-k-1} \ldots y_{t-k-N}$

und die direkten Eltern des Nachfolgers der Variablen sind:

$y_{t-1} \ldots y_{t-k-N+}$.

Aus den theoretischen Grundlagen ist bekannt, daß eine Variable unabhängig von einer anderen Variablen dieses Netzwerkes ist, wenn die Variablen innerhalb des Markov blankets bekannt sind. Deshalb wird die benötigte bedingte Dichte aus Gleichung (5) wie folgt bestimmt:

$$P(y^u \mid y^m) \propto P(y_{t-1} \mid y_{t-2}, \ldots, y_{t-1-N})$$
$$\times P(y_{t-2} \mid y_{t-3}, \ldots, y_{t-k}, \ldots, y_{t-2-N}) \ldots P(y_{t-k} \mid y_{t-k-1}, \ldots, y_{t-k-N}) \quad (5b).$$

Der hier beschriebene Fall eines fehlenden Meßwertes kann auch ohne Beschränkung der Erfindung auf mehrere nebeneinanderliegende fehlende Meßwerte ausgedehnt werden. Falls dies der Fall ist, muß beispielsweise nach dem erfindungsgemäßen Verfahren zunächst der eine Wert anhand seiner Nachbarn und Eltern und Vorfahren bestimmt werden und mit diesem bestimmten Wert dann der weitere Wert ermittelt werden. Dies kann solange hin- und hergehen bis eine hinreichende Genauigkeit erreicht wird. Für diesen Fall gilt:

$$y^u \subseteq \{y_{t-1}, y_{t-2}, \ldots, y_{t-n}\} \quad (5c)$$

Für alle fehlenden Werte der Zeitreihe zwischen dem Zeitpunkt $t-1$ und $t-N$, wobei weiterhin gilt:

$$y^m \subseteq \{y_{t-1}, y_{t-2}, \ldots, y_1\} \quad (5d),$$

welche die Zahl aller Meßwerte bis zum Zeitpunkt $t-1$ repräsentiert. Auch gilt

$$P(y^u \mid y^m) \propto P(y_{t-1}, \ldots, y_2, y_1) \quad (5e),$$

wobei die rechte Seite in (5e) aus Gleichung (4) erhalten wird. Im allgemeinen sind diese Integrale in den voranstehenden Gleichungen für die Funktion f(), falls dies eine nichtlineare Funktion ist, nicht analytisch lösbar. Details für die numerische Lösung mit Hilfe der Ziehung von Monte Carlo Proben werden im Zusammenhang mit Fig. 3 angegeben.

Fig. 3 zeigt zwei Repräsentationen eines neuronalen Netzes NN1 und NN2. Zum einen kann dies so aufgefaßt werden, daß NN1 zum einen Zeitpunkt vorliegt und NN2 zu einem anderen Zeitpunkt, es kann sich aber auch um zwei komplett verschiedene neuronale Netze handeln. Es sind zwei Datenleitungen dargestellt, mit welchen die neuronalen Netze kommunizieren können, im Falle einer Identität der beiden Netze sind damit zeitlich aufeinander folgende Austauschvorgänge gemeint. NN2 gibt Daten über die Datenstrecke 100 an NN1 weiter. NN1 gibt Daten über die Datenstrecke 150 an NN2 weiter. Die einzelnen Werte der Zeitreihe, welche in Fig. 1 und Fig. 2 dargestellt sind, sind der Einfachheit halber hier nicht mehr dargestellt. Es ist zu beachten, daß jedoch weiterhin die Voraussetzung für die anderen Figuren gelten.

Beim Training des Netzes mit fehlenden Daten gelten beispielsweise folgende Zusammenhänge: Für den Fall, daß $y_1, \ldots y_t$ mögliche Werte der Zeitreihe darstellen sollen $y^m \subseteq \{y_1, \ldots, y_t\}$ alle Meßwerte bezeichnen und $y^u = \{y_1, \ldots, y_t\} \setminus y^m$ alle unbekannten Werte bezeichnen. Das neuronale Netz, welches die Funktion f modelliert, werde beispielsweise mit einem Satz von Gewichten w parametrisiert. Dann gilt:

$$f(y_{t-1}, \ldots, y_{t-N}) \approx NN_w(y_{t-1}, \ldots, y_{t-N})$$

Es kann ohne Beschränkung der Erfindung jedoch auch ein anderer bekannter parametrisierbarer Funktionsapproximator verwendet werden. Die logarithmische Wahrscheinlichkeitsfunktion, auch log Likelihood Funktion genannt, lautet dann:

$$L = \log \int P^M(y_t, y_{t-1}, \ldots, y_2, y_1) dy^u$$

wobei dann die gemeinsame Wahrscheinlichkeitsdichte sich zu

$$P^M(y_t, y_{t-1}, \ldots, y_2, y_1) = P^M(y_N, \ldots, y_1) \prod_{l=N+1}^{t} P^M(y_l \mid y_{l-1}, \ldots, y_{l-N}) \quad (6)$$

approximiert und für das neuronale Netz folgender Zusammenhang für die Berechnung der Fehlerverteilungsdichte gilt:

$$P^M(y^t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-N}) = P_\varepsilon(y_t - NN_w(y_{t-1}, y_{t-2}, \ldots, y_{t-k}) \quad (7)$$

Für das Lernen mit Hilfe von Backpropagation, oder anderer gradientenbasierter Lernalgorithmen wird nun noch der Gradient der l garithmischen Wahrscheinlichkeitsfunkti n benötigt, welcher sich zu:

$$\frac{\partial L}{\partial w} = \sum_{l=N+1}^{t} \int \frac{\partial \log P^M(y_l | y_{l-1}, \ldots, y_{l-N})}{\partial w} P^M(y^{u(l)} | y^m) dy^{u(l)} \quad (8)$$

ergibt. Es ist anzumerken, daß hierbei von bekannten Ausgangsbedingungen für $y_1, \ldots, y_N$ ausgegangen wird. Für den Fall, daß eine Gaußverteilung für die Fehlerverteilung vorliegt ergibt sich daraus:

$$\frac{\partial L}{\partial w} \propto \sum_{l=N+1}^{t} \int (y_l - NN_w(y_{l-1}, \ldots, y_{l-N})) \frac{\partial NN_w(y_{l-1}, \ldots, y_{l-N})}{\partial w} P^M(y^{u(l)} | y^m) dy^{u(l)} \quad (8a)$$

wobei $y^{u(l)} = y^u \cap \{y_l, \ldots y_{l-N}\}$ die fehlenden Werte für die Eingänge des Netzwerkes darstellen und (8a) zeigt, daß falls alle $y_l \ldots y_{l-N}$ bekannt sind, das Integral verschwindet.

Falls die Meßwerte von einem zusätzlichen Rauschen überlagert werden ergeben sich die folgenden Zusammenhänge. Beispielsweise gilt wieder:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-N}) + \varepsilon_t$$

In dieser Variante der Erfindung soll jedoch kein direkter Zugriff auf y_t bestehen. Anstatt dessen wird die Zeitreihe

$$z_t = y_t + \delta_t$$

gemessen. Darin bedeutet δ_t ein unabhängiges Rauschen mit Mittelwert Null. Unter der Voraussetzung, daß $z = \{z_1 \ldots z_{t-1}\}$ und $y = \{y_1 \ldots y_t\}$ gelten, ergibt sich die Gesamtwahrscheinlichkeitsdichte zu:

$$P(y, z) = P(y_N, \ldots, y_1) \prod_{l=N+1}^{t} P(y_l | y_{l-1}, \ldots, y_{l-N}) \prod_{l=1}^{t} P(z_l | y_l) \quad (8b)$$

damit läßt sich die Rechenvorschrift für den erwarteten nächsten Wert der Zeitreihe angeben

$$E(y_t | z = \int f(y_{t-1}, \ldots, y_{t-N}) P(y_{t-1}, \ldots, y_{t-N} | z) dy_{t-1} \ldots dy_{t-N} \quad (9)$$

Ebenso kann der Gradient der Wahrscheinlichkeitsfunktion für das Training berechnet werden. Für den Fall, daß eine Gaußverteilung des Rauschens mit

$$z = \{z_1 \ldots z_t\}$$

vorliegt, ergibt sich:

$$\frac{\partial L}{\partial w} \propto \sum_{l=N+1}^{t} \int (y_l - NN_w(y_{l-1}, \ldots, y_{l-N})) \frac{\partial NN_w(y_{l-1}, \ldots, y_{l-N})}{\partial w}$$
$$\times P^M(y_1, \ldots, y_{l-N} | z) dy_1 \ldots dy_{t-N} \quad (9a)$$

Dem neuronalen Netz werden bei einer Variante des erfindungsgemäßen Verfahrens beispielsweise Werte zugeführt, die verrauscht oder nicht genau bestimmbar sind. Durch die Approximation der Gewichte im neuronalen Netz werden dabei über die Funktion f, welche dabei durch das neuronale Netz nachgebildet wird, neue Werte der Zeitreihe bestimmbar. Diese neuen Werte der Zeitreihe werden im Anschluß über die Datenstrecke 150 dem weiteren neuronalen Netz NN2 zugeführt, welches daraus wiederum durch Nachbildung der Funktion f neue Werte der Zeitreihe bestimmt. Dieser iterative Vorgang wird solange fortgesetzt, bis eine hinreichende Genauigkeit der zu bestimmenden Werte erreicht wird.

Zur genauen Bestimmung fehlender Werte mit Hilfe der Monte Carlo Methode wird von folgenden Grundlagen ausgegangen. Es ist hier zu beachten, daß alle Lösungen die Form

$$\int h(u, m) P(u | m) du \quad (9b)$$

aufweisen, wobei u den Satz von unbekannten Variablen und m den Satz von bekannten Variablen bedeutet. Ein Integral dieser Form kann beispielsweise gelöst werden, indem Zufallsproben der unbekannten Variablen

gemäß P(u | m) gez gen werden. Beispielsweis werden diese Proben mit $u^1, \ldots, u^s$ bezeichnet. Daraus ergibt sich folgender Zusammenhang für die Annäherung:

$$\int h(u, m) P(u|m) du \approx \frac{1}{S} \sum_{s=1}^{S} h(u^s, m) . \qquad (9c)$$

Es ist zu beachten, daß in dieser Gleichung u den Wert y_{t-k}, welcher fehlt, entspricht. Mit dieser erfindungsgemäßen Lösung reduziert sich das Problem also darauf, aus P(u | m) Proben zu ziehen. Für den Fall, daß lediglich eine Variable fehlt, reduziert sich das Problem also auf das Probenziehen aus einer einvariablen Verteilung, welche mit Hilfe des "samplingimportance-resampling" oder anderen sampling-Techniken [1] getan werden kann.

Für den Fall, daß mehr als ein Meßwert fehlt, wird die Situation etwas komplizierter. Der Grund besteht darin, daß die unbekannten Variablen in der Regel voneinander abhängen und daß von der Verteilung aller unbekannten Variablen gezogen werden muß. Eine allgemeine Lösung dafür gibt das Gibbs-Sampling [1] an. Beim Gibbs-Sampling werden die unbekannten Variablen mit Zufallswerten, oder besser mit geschätzten Ausgangswerten initialisiert, welche beispielsweise aus den benachbarten Werten der fehlenden Werte abgeleitet werden können. Im Anschluß wird eine der unbekannten Variablen u_i ausgewählt und eine Probe von $P(u_i|m, u\backslash u_i)$ gezogen; dann wird u_i auf diesen Wert gesetzt. Im Anschluß wird die Prozedur für die nächste unbekannte Variable wiederholt usw. Abgesehen von beispielsweise den ersten Proben werden die Proben vorzugsweise mit der korrekten Fehlerverteilungsdichte gezogen. Dies bedeutet jedoch, daß für alle Unbekannten, die jemals in der Zeitreihe auftraten, Proben gezogen werden müssen. In der Praxis kann aber beispielsweise das Zeitfenster aus dem Proben gezogen werden auf eine vernünftige Größe beschränkt werden. Beispielsweise kann diese Größe der Größe des Markov blankets für die fehlenden Werte entsprechen. Hierbei ist zu beachten, daß für den Fall, daß zwischen zwei fehlenden Werten N aufeinanderfolgende Werte bekannt sind, die Kopplung zwischen den Unbekannten aufbricht und daß weitere Werte der Zeitreihe deshalb nicht berücksichtigt werden müssen.

Das Probenziehen für zukünftige Werte ist nach dem erfindungsgemäßen Verfahren besonders einfach. Es ist jedoch zu beachten, daß es so nicht für deterministische Systeme funktioniert. Durch die erfindungsgemäße Vorgehensweise wird eine besonders einfache Lösung für diese besonders kompliziert erscheinenden Sachverhalte gefunden. Für die Vorhersage der Werte der Zeitreihe werden die nach der entsprechenden Verteilung gezogenen Werte substituiert und die Prognosen davon gemittelt, welches nach dem erfindungsgemäßen Verfahren den fehlenden Wert ergibt. Beim Training des Netzes wird beispielsweise der Durchschnitt der Fehlergradienten gebildet, indem zu deren Berechnung die Werte der Zeitreihe verwendet werden, welche mit den Proben bestimmt wurden. Beim Ziehen der Proben nach der Monte-Carlo-Methode kann beispielsweise wie folgt vorgegangen werden.

Es sollen beispielsweise K-Schritte in der Zukunft der Zeitreihe vorhergesagt werden. Im Zusammenhang mit dem oben besprochenen Gleichungen bedeutet dies, daß die Werte $y_{t-1}, \ldots, y_{t-K+1}$ fehlen, und daß unter diesen Voraussetzungen y_t vorhergesagt werden soll. Unter diesen Voraussetzungen ist die Monte Carlo Methode sehr einfach. Zunächst muß beispielsweise von der Verteilung $P(y_{t-K+1} | y_{t-K}, \ldots, y_{t-K-N})$ eine Probe gezogen werden. Diese Probe wird mit y_{t-K+1}^s bezeichnet. Mit dieser Probe und den vorangegangenen Messungen wird beispielsweise eine weitere Probe y_{t-K+2} aus der Verteilung $P(y_{t-K+2} | Y_{t-K+1}^s, \ldots, y_{t-K+1-N}$ usw. bis jede Probe für jede Unbekannte erzeugt wurde. Wenn man diese Prozedur S-mal wiederholt, so erhält man

$$E(y_t, M_{t-1}) = \frac{1}{S} \sum_{s=1}^{S} f(y_{t-2}^s, y_{t-1}^s, \ldots y_{t-N}^s) \qquad (9d)$$

Die experimentellen Befunde haben gezeigt, daß das erfindungsgemäße Verfahren selbst bei wenigen Proben viel besser arbeitet als die bisher verwendeten numerischen Lösungsverfahren.

Literatur

[1] Bernardo, J. M., Smith, A.F.M. (1994) Bayesian Theory. Wiley & Sons, pp 349—355.
[2] Buntine, W. L. and Weigend, A.S. (1991) Bayesian Back-Propagnation. Complex systems, Vol. 5, pp. 605—643.
[3] Ghahramani, Z. and Jordan, M.I. (1994) Supervised Learning from Incomplete Data via an EM approach. In: Cowan, J. D. et al., eds., Advances in Neural Information Processing Systems 6, Morgan Kaufman.
[4] Tresp, V. Ahmed, S. and Neuneier, R. (1994). Training Neural Networks with Deficient Data. In: Cowan, J. D. et al, eds., Advances in Neural Information Processing Systems 6, Morgan Kaufman.

Pat ntansprüche

1. Verfahren zur Aufbereitung einer Eingangsgröße eines neuronalen Netzes mit folgenden Merkmal n:
   a) es wird eine Zeitreihe aus einer Menge von M ßwert n der variablen Eingangsgröß gebildet, indem die Eingangsgröße zu diskreten Zeitpunkten bestimmt wird;
   b) von einem den Meßwerten überlagerten unkorrelierten Rauschen ndlicher Varianz, das im zeitli-

chen Mittel Null ist, wird die statistische Rauschverteilung ermittelt und/oder vorgegeben;
c) ein fehlender Meßwert in der Zeitreihe wird als Fehlwert aufbereitet, indem mindestens aus einem dem Fehlwert in der Zeitreihe benachbarten Meßwert nach der bekannten statistischen Rauschverteilung dessen statistische Fehlwert-Rauschverteilung berechnet wird und der Fehlwert berechnet wird, indem gemäß der Fehlwert-Rauschverteilung mindestens zwei Monte-Carlo-Proben des Fehlwertes gezogen werden, welche den Fehlwert ersetzen.

2. Verfahren nach Anspruch 1, bei dem ein vorherzusagender Fehlwert in der Zeitreihe als Prognosewert bestimmt wird, indem für den Fehlwert mehrere Monte-Carlo-Proben gezogen und deren Prognosewerte bestimmt werden, und der Prognosewert als arithmetrisches Mittel aus allen über die Proben bestimmten Prognosewerte berechnet wird.

3. Verfahren nach Anspruch 1 oder 2, bei dem von zwei fehlenden und unmittelbar benachbarten Meßwerten der Zeitreihe zunächst der Erste aufbereitet wird und im Anschluß daran unter Zuhilfenahme des zuerst aufbereiteten der Zweite bestimmt wird.

4. Verfahren nach einem der Ansprüche 1 bis 3, welches mehrfach ausgeführt wird.

5. Verfahren nach einem der Ansprüche 1 bis 4, bei dem das neuronale Netz anhand der Zeitreihe mit dem durch diese repräsentierten Verhalten eines technischen Systems trainiert wird, wobei bei einem Lernschritt bei der Backpropagation die Lernschrittweite für die auf 1 normierten Eingangsgrößen des neuronalen Netzes zu 0,1 dividiert durch die Anzahl der gezogenen Monte-Carlo-Proben festgelegt wird.

6. Verfahren nach einem der Ansprüche 1 bis 5, bei dem als statistische Rauschverteilung eine Gaußverteilung verwendet wird.

7. Verfahren nach einem der Ansprüche 1 bis 6 mit einer Zeitreihe der Form:

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-N}) + \varepsilon_t \qquad (1a)$$

mit: ε_t: statistische Rauschverteilung
y: Meßwerte der Zeitreihe
y_t: vom neuronalen Netz vorherzusagender Fehlwert
bei der die Funktion f bekannt ist oder durch das neuronale Netz modelliert wird und die statistische Fehlerverteilungsdichte zu:

$$P_\varepsilon(y_{t-1} - f(y_{t-1}, y_{t-2}, \ldots, y_{t-N})) = P(y^{t-1} \mid y_{t-1}, y_{t-2}, \ldots, y_{t-N}) \qquad (2a)$$

bestimmt wird woraus die Monte-Carlo-Proben $y_{t-k}{}^1, \ldots, y_{t-k}{}^s$ gezogen werden und sich ein erwarteter vom neuronalen Netz vorherzusagender Fehlwert mit den gezogenen Proben zu:

$$E(y_{t-k}, m) = \frac{1}{S}\sum_{s=1}^{S} f(y_{t-k}{}^s, m) \qquad (3a)$$

ergibt
mit: y_{t-k} fehlender Meßwert in der Zeitreihe $k \leq N$
m: alle bekannten Meßwerte der Zeitreihe
S: Anzahl der Proben.

8. Verfahren nach einem der Ansprüche 5 bis 7, bei dem das neuronale Netz mit mindestens einem aufbereiteten Wert nach folgender Lernfunktion trainiert wird:

$$w_{neu} = w_{alt} + \eta \frac{\partial L}{\partial w}$$

mit: w: Neuronengewichtung
L: logarithmische Wahrscheinlichkeitsfunktion
η: Lernfaktor
wobei gilt:

$$\frac{\partial L}{\partial w} = \frac{1}{S}\sum_{s=1}^{S}\sum_{l=N+1}^{t} \frac{\partial \log P^M(y^s_{l-1}, \ldots, y^s_{l-N})}{\partial w}$$

mit: NN_w: Werte der Funktion aus dem neuronalen Netz
wobei für y_l^s Meßwerte der Zeitreihe Verwendung finden und falls ein Wert nicht vorhanden ist aus der Wahrscheinlichkeitsverteilungsdichte:

$$P^M(y^l \mid y^m)$$

Monte-Carlo-Proben gez gen werden mit:
m: alle bekannten Meßwerte der Zeitreihe.

9. Verfahren nach Anspruch 6, bei dem das neuronale Netz mit mindestens einem aufbereiteten Wert nach folgender Lernfunktion trainiert wird:

$$w_{neu} = w_{alt} + \eta \frac{\partial L}{\partial w}$$

wobei gilt:

$$\frac{\partial L}{\partial w} = \frac{1}{S} \sum_{s=1}^{S} \sum_{l=N+1}^{t} y_l^s - NN_w(y_{l-1}^s, \ldots, y_{l-N}^s)) \frac{\partial NN_w(y_{l-1}^s, \ldots, y_{l-N}^s)}{\partial w}.$$

mit: NN_w: Werte der Funktion aus dem neuronalen Netz wobei für y_l^s Meßwerte der Zeitreihe Verwendung finden und, falls ein Wert nicht vorhanden ist, aus der Wahrscheinlichkeitsverteilungsdichte:

$$P^M(y^l \mid y^m)$$

Monte-Carlo-Proben gezogen werden mit:
m: alle bekannten Meßwerte der Zeitreihe.

10. Verfahren nach Anspruch 7, bei dem die statistische Rauschverteilung der Meßwerte nicht bekannt ist und die Meßwerte von einem weiteren Rauschen überlagert werden, dessen statistische Rauschverteilung bekannt ist, oder vorgegeben wird, mit einer Zeitreihe der Form:

$$z_t - y_t + \delta = f(y_{t-1}, y_{t-2}, \ldots, y_{t-N}) + \varepsilon_t \quad (1a^*)$$

mit: ε_t: unbekannte statistische Rauschverteilung
δ_t: bekannte statistische Rauschverteilung
y: Meßwerte der Zeitreihe
y_t: vom neuronalen Netz vorherzusagender Fehlwert
bei der die statistische Fehlerverteilungsdichte zu:

$$P_\varepsilon(y_{t-1}, y_{t-2}, \ldots, y_{t-N})) = P(y^t \mid y_{t-1}, y_{t-2}, \ldots, y_{t-N}) \quad (2a^*)$$

bestimmt wird und die Gesamtwahrscheinlichkeitsdichte über die Zeitreihe mit:

$$P(y,z) = P(y_1, \ldots, y_N) \prod_{l=N+1}^{t} P(y_l \mid y_{l-1}, \ldots, y_{l-N}) \prod_{l=1}^{t} P(z_l \mid y_l) \quad (3a^*)$$

bestimmt wird, so daß sich ein erwarteter vom neuronalen Netz vorherzusagender Fehlwert aus mindestens einem aufbereiteten Wert zu:

$$E(y_t \mid z) = \int f(y_{t-1}, \ldots, y_{t-N}) P(y_{t-1}, \ldots, y_{t-N} \mid z)) dy_{t-1}, \ldots, dy_{t-N} \quad (4^*)$$

ergibt, wobei für

$$P(y_{t-1}, \ldots, y_{t-N} \mid z)$$

entsprechend Monte-Carlo-Proben gezogen werden.

11. Verfahren nach den Ansprüchen 10 und 6, bei dem das neuronale Netz mit folgender Lernregel trainiert wird:

$$w_{neu} = w_{alt} + \eta \frac{\partial L}{\partial w}$$

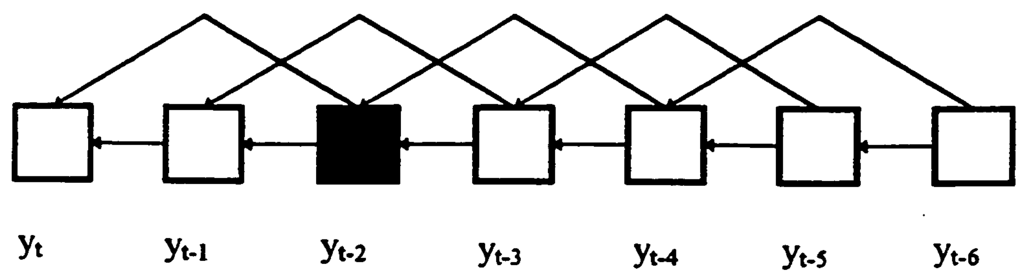
mit: w: Neuronengewichtung
L: logarithmische Wahrscheinlichkeitsfunkti n
η: Lernfaktor
und

$$\frac{\partial L}{\partial w} = \sum_{t=N+1}^{t} \int y_t - NN_w(y_{t-1},\ldots,y_{t-N}) \frac{\partial NN_w(y_{t-1},\ldots,y_{t-N})}{\partial w} \times P^M(y_{t-1},\ldots,y_{t-N}|z)dy_{t-1}\ldots dy_{t-N}$$

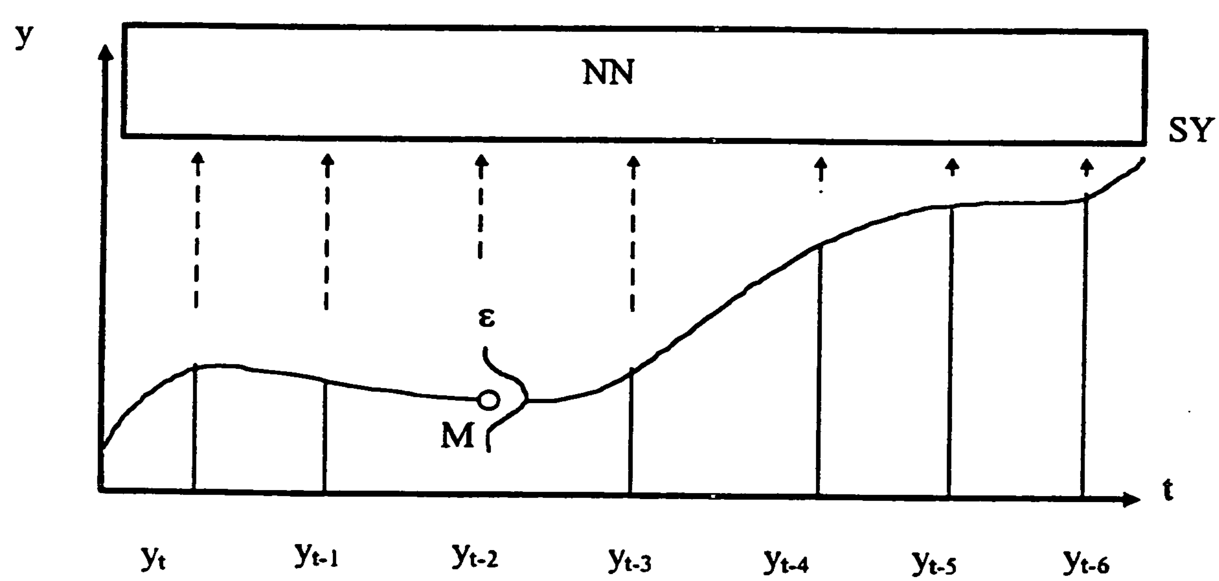
Hierzu 2 Seite(n) Zeichnungen

– Leerseite –

Figur 1



Figur 2



Nummer: DE 195 30 847 C1
Int. Cl.⁶: G 06 F 17/18
Veröffentlichungstag: 23. Januar 1997

Figur 3